APOBEC3G generates nonsense mutations in human T-cell leukemia virus type 1 proviral genomes in vivo.

AUTHOR(S):

Fan, Jun; Ma, Guangyong; Nosaka, Kisato; Tanabe, Junko; Satou, Yorifumi; Koito, Atsushi; Wain-Hobson, Simon; Vartanian, Jean-Pierre; Matsuoka, Masao

CITATION:

Fan, Jun ...[et al]. APOBEC3G generates nonsense mutations in human T-cell leukemia virus type 1 proviral genomes in vivo.. Journal of virology 2010, 84(14): 7278-7287

ISSUE DATE:

2010-07

URL:

http://hdl.handle.net/2433/126829

APOBEC3G Generates Nonsense Mutations in HTLV-1 Proviral Genomes *In Vivo*

Running title: Nonsense mutations by hA3G in HTLV-1

Jun Fan[1], Guangyong Ma[1], Kisato Nosaka[2], Junko Tanabe[1], Yorifumi Satou[1], Atsushi Koito[3], Simon Wain-Hobson[4], Jean-Pierre Vartanian[4] and Masao Matsuoka[1*]

[1]*Laboratory of Virus Control, Institute for Virus Research, Kyoto University, 53 Shogoin Kawahara-cho, Sakyo-ku, Kyoto 606-8507, Japan*

[2]*Department of Hematology and Infectious Diseases, Kumamoto University School of Medicine, 1-1-1 Honjo, Kumamoto 860-0811, Japan*

[3]*Department of Retrovirology and Self-Defense, Faculty of Medical and Pharmaceutical Sciences, Kumamoto University School of Medicine, 1-1-1 Honjo, Kumamoto 860-0811, Japan*

[4]*Molecular Retrovirology Unit, Institut Pasteur, 28 rue du Dr Roux 75724 Paris cedex 15, France*

Word count for abstract: 232

Word count for text: 5012

[*]Corresponding author. Mailing address for M. Matsuoka: Institute for Virus Research, Kyoto University, 53 Shogoin Kawahara-cho, Sakyo-ku, Kyoto 606-8507, Japan. Phone: +81-75-751-4048, FAX: +81-75-751-4049. E-mail: mmatsuok@virus.kyoto-u.ac.jp

ABSTRACT

Human T-cell leukemia virus type 1 (HTLV-1) induces cell proliferation after infection, leading to efficient transmission by cell-to-cell contact. After a long latent period, a fraction of carriers develop adult T-cell leukemia (ATL). In ATL cells, genetic changes in the *tax* gene were reported in about 10% of ATL cases. To determine genetic changes that may occur throughout the provirus, we determined the whole sequences of HTLV-1 provirus in 60 ATL cases. Abortive genetic changes including deletion, insertion and nonsense mutations were frequent in all viral genes except in the *HBZ* gene, which is transcribed from the minus strand of the virus. G-to-A base substitutions were the most frequent mutations in ATL cells. Sequence context of G-to-A mutations was in accordance with the preferred target sequence of human APOBEC3G (hA3G). Target sequences of hA3G were less frequent in the plus strand of the *HBZ* coding region than in other coding regions of HTLV-1 provirus. Nonsense mutations in viral genes including *tax* were also observed in proviruses from asymptomatic carriers, indicating that these mutations were generated during reverse transcription and prior to oncogenesis. That hA3G targets the minus strand during reverse transcription explains why the *HBZ* gene, which is encoded by the minus strand of provirus, is not susceptible to such nonsense mutations. HTLV-1 infected cells likely take advantage of hA3G to escape from the host immune system by losing expression of viral proteins.

INTRODUCTION

Human T-cell leukemia virus type 1 (HTLV-1) is the causative agent of both adult T-cell leukemia (ATL) and inflammatory HTLV-1 associated myelopathy/tropical spastic paraparesis (HAM/TSP) (9, 40). HTLV-1 is a complex retrovirus that encodes regulatory genes, *tax* and *rex*, and accessory genes, *p12*, *p13*, and *p30* (6, 29). An accessory gene, *HTLV-1 bZIP factor* (*HBZ*), is transcribed from the 3' long terminal repeat (LTR) as an anti-sense transcript (10, 47). Previous studies have shown that Tax expression can immortalize T-lymphocytes *in vitro* (1, 11), and *in vivo* the expression of Tax in transgenic mice causes various tumors depending on the tissue-specific promoter that expresses Tax (12, 14, 20). In more than half of human ATLs, the *tax* gene is not transcribed (41). Three mechanisms have been identified to inactivate Tax expression (29): 1) abortive genetic changes of the *tax* gene (7, 41), 2) silencing by DNA methylation in the 5'LTR (18, 43), and 3) deletion of the 5'LTR (31, 42). Since Tax is a major target of cytotoxic T-lymphocytes (CTLs) *in vivo* (17), the tumor cells might escape from the host immune system by suppressing Tax expression. However, the actual mechanism used to create genetic changes in the *tax* gene remains to be elucidated. Nonsense mutations in the *tax* gene have also been observed in HTLV-1 carriers and the mechanism to generate these mutations is similarly unknown (8).

As a host defense against retroviruses, mammalian cells employ the APOBEC3 family, which causes deamination during reverse transcription, resulting in nucleotide mutations (3, 26). Human APOBEC3G (hA3G) deaminates cytosine residues of single-stranded DNA during reverse transcription, resulting in high levels of plus-strand

G-to-A mutations. The human immunodeficiency virus (HIV) nucleocapsid is critical for incorporation of hA3G into virions (2, 24), while its accessory protein, Vif, counteracts hA3G. Vif inhibits hA3G packaging into HIV-1 virion through ubiquitination and proteasomal degradation of hA3G (39, 49). HTLV-1 does not encode a protein analogous to Vif that inactivates hA3G. Instead, it has been demonstrated that a domain of the HTLV-1 nucleocapsid suppresses the incorporation of hA3G into the virion (4). Consistent with this finding, it has been reported that G-to-A mutations are rare in HTLV-1 carriers (25).

In this study, we analyzed the sequences of whole proviruses in ATL and carrier cells, and found that most nonsense mutations in the proviruses were caused by deamination. The sites of nonsense mutation coincided with the preferred target sequences of hA3G. In ATL cases, non-sense mutations, deletions and insertions were detected in most of the viral genes except the *HBZ* gene, supporting the critical role of the *HBZ* gene in ATL.

MATERIALS AND METHODS

Cell lines and clinical samples. MT-2 and MT-4 are HTLV-1-transformed cell lines. MT-1 and TL-Om1 were derived from leukemia cells. Jurkat cells were negative for HTLV-1. These cells were cultured in RPMI-1640 medium supplemented with 10% fetal bovine serum, 100 U/ml penicillin, 100 μg/ml streptomycin and 2 mM L-glutamine at 37˚C under a 5% CO_2 atmosphere. Clinical samples were collected from 60 ATL patients and 10 HTLV-1 carriers. Genomic DNA was extracted from peripheral blood mononuclear cells using standard phenol-chloroform methods. This study was conducted according to the principles expressed in the Declaration of Helsinki. The study was approved by the Institutional Review Board of Kyoto University (G204). All patients provided written informed consent for the collection of samples and subsequent analysis.

Sequencing of complete provirus genomes in ATL cells. The copy number of HTLV-1 provirus in ATL cases was determined by inverse PCR (5) and ATL samples with one complete provirus were selected for direct sequencing. The complete provirus genome was amplified as two halves (fragment I and II) from genomic DNA samples derived from ATL patients. Fragment I was amplified using primers 5'-TGACAATGACCATGAGCCCCAAATATCC-3' and 5'-CGGCTATTAAGACCAGGAAG-3'. PCR conditions were as follows: 1 cycle of 94^{o}C for 5 min, 30 cycles of 94^{o}C for 30 s, 64 ^{o}C for 30 s and 72 ^{o}C for 5 min, 1 cycle of 72 ^{o}C for 10 min. Fragment II was amplified using primers 5'-

AGAAACAAGCTCAGAAGCTA -3' and 5'-TGTACTAAATTTCTCTCCTGAGAGTGC-3'. PCR conditions were as follows: 1 cycle of 94°C for 5 min, 30 cycles of 94°C for 30 s, 60 °C for 30 s and 72 °C for 5 min, 1 cycle of 72 °C for 10 min. PCR products were then subjected to nested PCR to amplify I-1 and I-1 from fragment I, and II-1 and II-2 from fragment II. Primers for I-1 were 5'-TGACAATGACCATGAGCCCCAAATATCC-3' and 5'-GAGCTTAAAGTGATCTTGG-3'. Primers for I-2 were 5'-TTCCGATAGCCTTGTTCTCA-3' and 5'-CGGCTATTAAGACCAGGAAG-3'. Primers for II-1 were 5'-TGGTATTATTTCAAGCTTCC-3' and 5'-AAATGCAGGAGTTGGGGATT-3'. Primers for II-2 were 5'-AACTGTCTAGTATAGCCATC-3' and 5'-TGTACTAAATTTCTCTCCTGAGAGTGC-3'. PCR conditions for the nested PCR were as follows: 1 cycle of 94°C for 5 min, 28 cycles of 94°C for 30 s, X °C for 30 s and 72 °C for 2.5 min, 1 cycle of 72 °C for 10 min. X was 60 °C for I-1 and II-2, 58 °C for I-2 and II-1. The final PCR products were gel purified and sequence was determined by direct sequencing.

Sequencing of the *pol* and *tax* regions in HTLV-1 carriers. Proviral load was determined as previously reported (32). Herculase II Fusion DNA Polymerase (Stratagene, La Jolla, CA), a Pfu-based DNA polymerase, was used to amplify *pol* and *tax* regions from genomic DNA derived from HTLV-1 carriers. Primers to amplify the *pol* region were 5'- TACACCTTGCAATCCTATGG-3' and 5'-

GCTAGGCTTGCCTAGATGGG-3'. PCR conditions were as follow: 1 cycle of 95°C for 5 min, 30 cycles of 95°C for 20 s, 60 °C for 20 s and 72 °C for 50 sec, 1 cycle of 72 °C for 5 min. Primers to amplify the *tax* region were 5'-GGTCCTCCGGGCATGACACA-3' and 5'- TCTCCACGCTTTTATAGACT-3'. PCR condition were as follow: 1 cycle of 95°C for 5 min, 28 cycles of 95°C for 20 s, 68 °C for 20 s and 72 °C for 50 sec, 1 cycle of 72 °C for 5 min. PCR products were purified, cloned and at least 20 clones sequenced. DNA polymerase fidelity was studied by using a clone carrying an amplified *tax* region as template instead of genomic DNA. The *tax* region was amplified using the same PCR conditions described above, cloned and 25 clones were sequenced.

Sequencing of HTLV-1 genome in non-tumor cells derived from ATL samples. ATL patient-derived tumor cells are considered a mixture of a major population of monoclonal expanded tumor cells and a minor population of HTLV-1 infected non-tumor cells. To amplify the HTLV-1 genome from non-tumor infected cells, ATL samples carrying defective HTLV-1 genome based on direct sequencing were selected. Target regions were amplified by setting one of the primers at the deleted region. Sequences were determined by direct sequencing.

Mutation analysis. Nucleotide sequences of all ATL samples were aligned using GENETYX-MAC Ver.13.0 software. Mutations were determined after filtering for i) polymorphisms, defined as a nucleotide substitution occurring in >5 cases; ii) linkers,

defined as >7 common nucleotide substitutions occurring in >2 cases. With these criteria, the sequence of an ATL case, ATL-25, showed the fewest mutations (3 single nucleotide substitutions with no deletions or insertions) and was selected as a standard sequence. The nucleotide positions of identified mutations were numbered relative to this standard sequence, which has been submitted to the DDBJ (DNA Data Bank of Japan) database under the accession number AB513134.

For carrier samples, the sequence of a given region from all clones amplified from a carrier was aligned and compared with the standard sequence described above with mutated bases corrected. When a nucleotide substitution occurred in all clones, it was considered a polymorphism and excluded from the mutation repertoire.

cDNA synthesis and quantitative real-time RT-PCR. Total RNA was extracted from transfectants using Trizol reagent (Invitrogen, Carlsbad, CA). RNA was treated with DNase I (Invitrogen) to eliminate the genomic DNA. cDNA was synthesized from 1 μg of total RNA with the Superscript preamplification system (Invitrogen) according to the manufacturer's protocol. cDNA product was quantified by real-time RT-PCR with Power SYBR Green PCR Master Mix and 7900HT Fast Real Time PCR System (Applied Biosystems) according to the manufacturer's instructions. Specific primers for *hA3G* gene were 5'-CGCGTGCCACCATGAAGATC-3' (forward) and 5'-TGTGGGTGGATCCATCGAGT-3' (reverse). Specific primers for the human *AID* gene were 5'-TTCACCGCGCGCCTCTACTT-3' (forward) and 5'-GCTGTCTGGAGAGACGAACT-3' (reverse). Target cDNA was normalized to the

amount of endogenous mRNA of *ACTB*. Primers used for *ACTB* were 5'-AGGCCAACCGCGAGAAGATG-3' (forward) and 5'-CCAGAGGCGTACAGGGATAG-3' (reverse). PCRs were carried out in triplicate. Data was analyzed by the comparative C_t method according to the manufacturer's protocols.

***In vitro* editing of HTLV-1 genome by hA3G.** QT6 quail cells were transfected with an infectious molecular clone of HTLV-1 (pX1MT-M) (30) in the presence or absence of hA3G-expression plasmid. Twenty-four hours post-transfection, HTLV-1-producing QT6 cells were treated with 200 μg/ml of mitomycin C at 37˚C for 30 min, well-washed and then co-cultivated 1:1 with fresh QT6 cells (13). 72 hours later, cells were collected and genomic DNA was extracted. hA3G-mediated DNA hypermutation were detected using 3DPCR that allows differential amplification of G-to-A hypermutants using a Taq DNA polymerase as previously described (25). The first round of PCR was performed with a denaturation temperature of 94°C using the primer pair 5'-CTGCAGATACAAAGTTAACC-3' (forward) and 5'-TGGAGGAAGGAGGGTGGAAT-3' (reverse). PCR conditions were 94˚C for 5 min, followed by 35 cycles of 94˚C for 30 sec, 58˚C for 30 sec and 72˚C for 30 sec, with a final extension for 10 min at 72˚C. Then, 0.5 μl of the first-round reaction was used as template for the second round of PCR using the following primers: 5'-CCATGCTTATTATCAGCCCA-3' (forward) and 5'-GTTGGGGGTTGTATGAGTGA-3' (reverse). PCR program was the same as the

first round of reaction except that the denaturation gradient ranged from 94-82˚C. PCR products were purified, cloned and up to 20 clones were sequenced.

RESULTS

HTLV-1 proviral sequence changes in ATL. Previous studies showed that Tax expression was frequently lost by genetic changes in the *tax* gene, by DNA methylation of the 5'LTR, and by deletion of the 5'LTR (29). Similarly, although not well studied to date, the expression of other viral proteins might also be lost through genetic changes in the provirus sequence. To analyze the genetic changes of the HTLV-1 provirus in ATL cells, we determined the whole sequence of proviruses (9034bp) in 60 ATL cases. For this study, we chose ATL cases with HTLV-1 proviruses containing both LTRs. As shown in Table 1, we detected genetic changes (nonsense mutations, deletions and insertions) in the coding region of viral genes in 28 of 60 ATL cases (46.7%). Deletions in the proviruses were detected in 27 cases (45%) while insertions were found in 10 cases (16.7%) (See Table S1 in the supplemental material). These deletions and insertions resulted in the loss or the truncation of protein(s) in 18 cases (Table 1 and see Table S1 in the supplemental material).

Nonsense mutations, deletions, and insertions in the viral genes were found in all viral genes except the *HBZ* gene (Table 1). Since *p30* and *p13* use an identical coding frame, the same mutations generate nonsense changes in both *p30* and *p13* (Fig. 1). The frequency of genetic changes in the *tax* gene was 16.7% (10/60) (Table 1), which was higher than in previous studies (7, 41). 27 of a total of 29 nonsense mutations preferentially accumulated in tryptophan codons (27/29: 89.7%) (See Table S2 in the supplemental material). Since the tryptophan codon is TGG, a G-to-A mutation generates either a TGA or a TAG stop codon. In one nonsense mutation, the codon for

arginine (CGA) was converted to the TGA stop codon by a C-to-T mutation. Interestingly, nonsense mutations tended to accumulate in the same cases (Table 1).

It was difficult to discriminate between mutations and polymorphisms in provirus from ATL patients since the proviral sequences in non-leukemic cells could not be analyzed in most ATL cases without an internal deletion of provirus. Some mutations might be characteristic of a subgroup of proviruses (polymorphism). Since nonsense mutations at the *tax* gene (7469) were detected in 5 ATL cases, we tentatively judged base substitutions as polymorphisms when more than 5 identical base substitutions were observed in different cases. Based on this criterion, we observed 591 mutations in all cases (See Table S3 in the supplemental material). Among them, the G-to-A mutation was the most frequent, contributing to 28.8% of the total mutations. Other frequently detected mutations were C-to-T (23.7%), A-to-G (17.1%) and T-to-C (12.9%). Frequent G-to-A and C-to-T mutations strongly suggested the activity of deamination enzymes in the generation of these mutations.

Mutations in leukemic and non-leukemic cells in ATL cases. In ATL patients, non-leukemic HTLV-1 infected cells co-exist with leukemic cells. To analyze the proviral sequences of non-leukemic cells, we amplified proviral sequences using primers within regions deleted in leukemic cells to avoid amplification of provirus from leukemic cells. Nonsense mutations that were observed in leukemic cells were not detected in non-leukemic cells (Fig. 2), indicating that nonsense mutations generated in the provirus of leukemic cells were selected for. We compared the whole sequences of

HTLV-1 provirus in both leukemic cells and non-leukemic cells from five ATL cases. As shown in Table 2, only 37% of base substitutions were considered to be real mutations since the nucleotide differed between leukemic cells and non-leukemic cells. However, 71% of G-to-A transitions were actually mutations while 73% of other base substitutions were polymorphisms (Table 2), indicating that G-to-A mutations are predominant in ATL cells.

Association of G-to-A mutations with human APOBEC3G. A high frequency of G-to-A mutations suggests the role of deaminase(s) in generating these mutations. Deamination enzymes are known to have individually preferred target sequences for deamination (44). Sequences surrounding the nonsense mutations generated by the G-to-A mutation in HTLV-1 proviruses showed a predominance of the 5'-GG dinucleotide context (target underlined) (Fig. 3A). Among trinucleotides containing GG dinucleotides, CGG, TGG and GGG were preferred (Fig. 3A); these are consistent with the target sequences of hA3G (48). The AGG sequence was also targeted in HTLV-1, unlike in HIV-1. When we checked tetrameric sequences that contain a central GG dinucleotide, CGGG and TGGG were the preferred targets in the HTLV-1 provirus, similar to those reported for HIV-1 (Fig. 3B). As in HIV-1, a C at the +2 position (NGGC) was disfavored in HTLV-1. These findings suggest that the observed G-to-A mutations were generated by hA3G. Another deaminase, APOBEC3F (hA3F), which is largely coexpressed with hA3G, is also reported to target single-stranded minus DNA (46). Unlike hA3G, which favors the 5'-GG dinucleotide, the consensus target sequence of hA3F is 5'-GA (21). G-to-A mutations at GA sites contributed 13.4% of all G-to-A

mutations (Fig. 3A), suggesting that hA3F might also play a role in these mutations in the HTLV-1 provirus.

Since hA3G targets the minus strand during reverse transcription, nonsense mutations generated by G-to-A mutations are not found in the *HBZ* gene**.** A G-to-A mutation in the proviral sense strand caused a nonsense mutation in the *p12* gene (TGG to TGA), whereas the same mutation generated a missense mutation in the *HBZ* gene (CCA to TCA) (Fig. 1). G-to-A mutations occurred most frequently at GG sites with a T or C at their 5' terminus (Fig. 3A). We therefore counted the TGG or CGG sites within individual HTLV-1 regions (Table 3). As NGGC (N indicates T or C) was disfavored by hA3G (Fig. 3B), it was not included. Among HTLV-1 regions, the plus-strand sequence of the *HBZ* gene had the fewest TGG and CGG sites, even when the number of sites was normalized by size of the region. This indicates that mutations in the *HBZ* gene-coding region caused by hA3G are also rare.

Mutations in the HTLV-1 proviral sequences from carriers. Frequent detection of G-to-A mutations in HTLV-1 proviral sequences in ATL cases suggests that these nonsense mutations potentially occur in both HTLV-1 carriers and ATL patients. To explore this possibility, we next studied the proviral sequences (*pol, env, p12, p13/p30, p27, tax,* and *HBZ*) in asymptomatic HTLV-1 carriers. The proviral sequences in asymptomatic carriers were amplified using a high fidelity DNA polymerase with proofreading activity, and after subcloning, the sequences were determined for at least 20 subclones. As seen in ATL cases, the mutation frequency differed drastically among

individual carriers (representative data are shown in Table S4 and S5 in the supplemental material), and the distribution of base substitutions among clones varied within individual carriers. G-to-A mutations accounted for 81% and 72 % of all mutations in the *pol* and the *tax* genes, respectively (Table 4 and 5). High frequency of G-to-A mutations was also observed in other viral genes (data not shown). Analyses of the sequence context of G-to-A mutations showed a predominance of GG sequences, which is consistent with the finding that hA3G deaminates the viral genome during reverse transcription (48). In order to compare the frequencies of nonsense mutations in carriers versus ATL cases, the occurrence of nonsense mutations was further analyzed. Consistent with findings in ATL cases, nonsense mutations were detected in the *pol*, *env, p30/p13*, *p27* and *tax* genes, but not in the *p12* or *HBZ* genes (Table 6). As in ATL cases, nonsense mutations in carriers were most frequently observed in the *tax* and *pol* genes. This result suggests that there is no bias for mutations in specific viral genes in ATL cases compared with carriers. These results suggest that nonsense mutations in provirus are not generated during oncogenesis, but are present in the carrier state.

In addition to G-to-A mutation, C-to-T mutations were detected in the *pol* and *tax* genes in carrier cells as in ATL cells (See Table S2 in the supplemental material) as well as other viral regions (data not shown). Comparative analyses of leukemic cell and non-leukemic cells showed that 79% of C-to-T base substitutions were polymorphisms (Table 2). However, after G-to-T mutations, C-to-T mutations were still the most frequent in ATL patients and HTLV-1 carriers.

Expression of the *hA3G* gene and activation-induced deaminase gene

In order to analyze the correlation between G-to-A mutation and *hA3G* expression, we first studied the mRNA level of *hA3G* using real-time PCR assay. The *hA3G* gene was expressed in normal T cells and the expression level of *hA3G* moderately decreased in ATL cells. In ATL samples, there was no obvious correlation between the expression level of *hA3G* and number of G-to-A mutations in the proviral genome.

Although only 21% of C-to-T base substitutions were mutations (Table 2), C-to-T was the second most frequent mutation (See Table S2 in the supplemental material). Activation-induced deaminase (AID) was found to be a key factor for the switch recombination of immunoglobulin (Ig), and is also implicated in hypermutations of Ig genes (15). AID deaminates cytosine to uracil. Aberrant expression of AID driven by NF-κB activation is found in gastric cancers and hepatomas, and AID expression has been invoked to explain induced mutations observed in cancer cells (28). Therefore, we studied *AID* expression in 7 ATL cases with different proviral mutation frequencies. As shown in Fig.4, the *AID* gene transcription increased in ATL-15. In this case, we found six C-to-T mutations in the provirus, and the sequence context showed more AID preferred sequences in C-to-T mutated sites compared to other cases without AID activation (data not shown). Thus, AID might also play a role in the deamination of the HTLV-1 genome in at least some ATL cases. Since AID is thought to target double strand DNA of provirus, it likely generates C-to-T mutations in the plus strand of provirus. However, C-to-T base substitutions contained many polymorphism, as shown in Table 2. Therefore, correlation between C-to-T mutations and AID expression

remains to be studied.

***In vitro* editing of the HTLV-1 genome by hA3G**

Previous reports have suggested that HTLV-1 is relatively resistant to the antiviral effect of hA3G (4), consistent with our finding that the frequency of G-to-A mutations throughout the HTLV-1 proviral genome was low. In order to provide direct evidence that G-to-A mutations observed in the proviruses are the result of hA3G-mediated genome editing, we studied *in vitro* the editing effect of hA3G on the HTLV-1 genome. HTLV-1 viruses were generated in the presence of exogenous hA3G to allow packaging into budding viral particles, a step that is required for exerting editing activity of the cytidine deaminase. After a round of virus infection, the HTLV-1 DNA was analyzed using a highly sensitive PCR-based protocol, referred to as 3D PCR, that is capable of amplifying G-to-A hypermutated genomes as described in the Materials & Methods. Differential amplification of a tax region from cells infected with viral particles produced in the presence of hA3G was achieved when the denaturation PCR temperature was lowered to between 84-86˚C (data not shown). All sequences of differentially amplified PCR clones exhibited extensive G-to-A hypermutation, and the number (percentage) of G-to-A transitions ranged from 8 (13%) to 29 (48%) per clone (Fig. 5). In ATL cases in the same region, G-to-A mutations at four nucleotide positions resulted in premature stop codons in the *p30* and *tax* genes (Fig. 1). The occurrence of such mutations was observed *in vitro* when viruses were produced together with hA3G. G-to-A mutations at two sites, in which nonsense mutations were more frequently

detected in ATL cases, were also frequent among 3D PCR clones. As expected, a preference for GG dinucleotide context in hA3G-mediated G-to-A editing was observed (Fig. 5C). Our data suggested that hA3G is indeed able to actively deaminate HTLV-1 viral genome during reverse transcription although with low editing frequency, leading to G-to-A mutations in the plus-strand of the provirus.

DISCUSSION

In this study, in order to acquire a full mutation spectrum of HTLV-1 virus, sequences of proviral genomes were determined in 60 ATL cases as well as 10 asymptomatic carriers. In addition to deletion and insertion, we identified base substitutions, and found that G-to-A mutations were most prevalent. In addition to the observation that G-to-A mutations in ATL and carriers occurred at the preferred target sequence of hA3G, we also experimentally confirmed that hA3G actively introduced G-to-A mutations to the HTLV-1 viral genome. It has been reported that hypermutation in the HTLV-1 genome was a rare event (35), and HTLV-1 was relatively resistant to the antiviral activity of hA3G (33). Unlike HIV, which counteracts hA3G by producing Vif, HTLV-1 does not express an accessory protein that interferes with hA3G. A peptide motif in the C-terminus of the HTLV-1 nucleocapsid has been shown to inhibit hA3G packaging into nascent virions (4). However, inhibition of hA3G induced G-to-A mutations by HTLV-1 was partial (37). Previous studies analyzed small regions of provirus amplified by PCR in carriers or HAM/TSP patients, while proviral sequences of ATL cells were determined by direct sequencing. Since previous studies used error-prone Taq polymerase, it was very difficult to identify mutations after cloning (36). The number of ATL cases was so limited in the previous studies that mutations could not be discriminated from polymorphisms (34). This study first analyzed the whole sequence of 60 ATL cases by direct sequencing, and also determined the proviral sequences of clones derived from HTLV-1 carriers using a precise DNA polymerase, which enabled us to identify mutations in HTLV-1 genomes. Consistent with previous

studies (34, 35), our results showed that mutations were relatively rare. However, among the rare mutations in HTLV-1 genomes, G-to-A mutations in the target sequences of hA3G were predominant. In addition to G-to-A mutations, other mutations were also observed in ATL and carriers, indicating the role of other deaminases including AID and error-prone reverse transcriptase (27).

This study demonstrated important roles for cytidine deaminase(s) in generating mutations in HTLV-1 proviruses. So far, at least 12 cytidine deaminases have been identified, many of which share significant sequence homology. They are considered to have evolved through gene transpositions and duplications (45). Although some members, including APOBEC1 and AID, target cellular genes, most human APOBEC proteins are thought to defend the host against retroviruses. Our study suggests that hA3G is responsible for G-to-A mutations in HTLV-1. hA3G binds to single-stranded DNA, and preferentially deaminates CCCA or CCCG sequences during reverse transcription (48). This strategy of hA3G can induce nonsense mutations in the plus-strand coding sequence since TGG is a target of hA3G, resulting in nonsense mutations like TAG or TGA. However, the *HBZ* gene is encoded by the minus strand of the provirus. The *HBZ* gene is therefore much less susceptible to nonsense mutations generated by hA3G. Furthermore, there are few target sequences for hA3G in the plus strand of the *HBZ* coding RNA. In addition, the coding sequence of the *HBZ* gene overlaps that of *p12* gene. Thus, HBZ further avoids missense mutation in the minus strand.

The frequencies of G-to-A changes in the HTLV-1 provirus (0.21% for the *pol* gene

and 0.11% for the *tax* gene) in the carriers are slightly lower than that for vif positive HIV-1 (0.57%) (48), indicating that HTLV-1 is resistant to hA3G during reverse transcription. Why are the few nonsense mutations that do occur retained in the HTLV-1 provirus? The difference between HTLV-1 and HIV-1 is related to their distinct strategies of propagation. HIV-1 replicates vigorously *in vivo*, producing tremendous numbers of viral particles. Viruses with nonsense mutations cannot replicate, and thereby disappear *in vivo*. On the other hand, HTLV-1 promotes the proliferation of the infected cells themselves by the action of its regulatory and accessory genes. Therefore, HTLV-1 infected cells can proliferate despite nonsense mutations occurring in most of the viral genes, provided the cells retain the minimum set of viral genes that relate to proliferation. This might be a reason why so many nonsense mutations in various viral genes remain in the provirus. The only gene without nonsense mutation is *HBZ*. It is likely that the *HBZ* gene is indispensable for proliferation of ATL cells, as our previous studies reported (38).

Previous studies suggest that Tax is critical for proliferation of HTLV-1 infected cells and oncogenesis. However, we reported that Tax expression is frequently disrupted by three mechanisms (29). It has been speculated that Tax expression is not necessary in the late stages of ATL. Rather, since Tax is a major target of CTLs, ATL cells without Tax expression are selected during leukemogenesis. However, as shown in this study, nonsense mutations were likely generated by hA3G during reverse transcription , indicating that Tax expression was not necessary even in the carrier state. In one carrier, 9 of 21 clones shared a nonsense mutation in the *tax* gene, demonstrating clonal

expansion of HTLV-1 infected cells with this mutation (Table S5). Furthermore, 7 of 60 ATL cases contained nonsense mutations in the *tax* gene. HTLV-1 infected cells with nonsense mutations could proliferate *in vivo* and be transformed to ATL cells. A nonsense mutation in the *tax* gene has been detected in asymptomatic carriers (8). This study suggests that this nonsense mutation was generated by hA3G during reverse transcription. Another explanation is that loss of the expression of some viral gene(s) by nonsense mutations benefits the cells. Tax generates DNA damage that activates checkpoints (16, 22). In addition, it has been reported that HTLV-1-infected cells become cell cycle-arrested due to the activity of Tax (19, 23). Therefore, HTLV-1 infected cells with nonsense mutations in the tax gene might have growth-advantage by loosing Tax expression. We previously reported that HTLV-1 provirus lacking the 5'LTR and the second exon of the *tax* gene was detected in ATL cases. By sequencing the integration sites in such defective provirus, we found 6 bp short repeats generated by integrase. This finding showed that defective proviruses that could not produce Tax were generated before integration (31). Taken together, these findings indicate that Tax is not necessary for oncogenesis at least in some ATL cases.

This study suggests that nonsense mutations in the HTLV-1 provirus are generated by hA3G in both ATL cases and HTLV-1 carriers. That hA3G targets the minus strand during reverse transcription explains why the *HBZ* gene is not susceptible to such nonsense mutations. On the contrary, HTLV-1 infected cells take advantage of hA3G to escape from the host immune system by losing expression of other viral proteins, while the *HBZ* gene remains intact.

ACKNOWLEDGMENTS

We thank Kuan-Teh Jeang for helpful discussion, Aaron Coutts for his valuable suggestions and kind revision of the manuscript, and David Derse for providing pX1MT-M. This study was supported by a Grant-in-aid for Scientific Research from the Ministry of Education, Science, Sports, and Culture of Japan, a grant from the Naito Foundation, and a grant from the Sumitomo Foundation to M.M., a grant from Japan Leukemia Research Fund to Y.S. Y.S. is supported by JSPS Research Fellowships for Young Scientists.

REFERENCES

1. **Akagi, T., H. Ono, and K. Shimotohno.** 1995. Characterization of T cells immortalized by Tax1 of human T-cell leukemia virus type 1. Blood **86:**4243-9.
2. **Cen, S., F. Guo, M. Niu, J. Saadatmand, J. Deflassieux, and L. Kleiman.** 2004. The interaction between HIV-1 Gag and APOBEC3G. J Biol Chem **279:**33177-84.
3. **Chiu, Y. L., and W. C. Greene.** 2008. The APOBEC3 cytidine deaminases: an innate defensive network opposing exogenous retroviruses and endogenous retroelements. Annu Rev Immunol **26:**317-53.
4. **Derse, D., S. A. Hill, G. Princler, P. Lloyd, and G. Heidecker.** 2007. Resistance of human T cell leukemia virus type 1 to APOBEC3G restriction is mediated by elements in nucleocapsid. Proc Natl Acad Sci U S A **104:**2915-20.
5. **Etoh, K., S. Tamiya, K. Yamaguchi, A. Okayama, H. Tsubouchi, T. Ideta, N. Mueller, K. Takatsuki, and M. Matsuoka.** 1997. Persistent clonal proliferation of human T-lymphotropic virus type I-infected cells in vivo. Cancer Res **57:**4862-7.
6. **Franchini, G., R. Fukumoto, and J. R. Fullen.** 2003. T-cell control by human

T-cell leukemia/lymphoma virus type 1. Int J Hematol **78:**280-96.

7. **Furukawa, Y., R. Kubota, M. Tara, S. Izumo, and M. Osame.** 2001. Existence of escape mutant in HTLV-I tax during the development of adult T-cell leukemia. Blood **97:**987-93.

8. **Furukawa, Y., M. Tara, S. Izumo, K. Arimura, and M. Osame.** 2006. HTLV-I viral escape and host genetic changes in the development of adult T cell leukemia. Int J Cancer **118:**381-7.

9. **Gallo, R. C.** 2005. The discovery of the first human retrovirus: HTLV-1 and HTLV-2. Retrovirology **2:**17.

10. **Gaudray, G., F. Gachon, J. Basbous, M. Biard-Piechaczyk, C. Devaux, and J. M. Mesnard.** 2002. The complementary strand of the human T-cell leukemia virus type 1 RNA genome encodes a bZIP transcription factor that down-regulates viral transcription. J Virol **76:**12813-22.

11. **Grassmann, R., C. Dengler, I. Muller-Fleckenstein, B. Fleckenstein, K. McGuire, M. C. Dokhelar, J. G. Sodroski, and W. A. Haseltine.** 1989. Transformation to continuous growth of primary human T lymphocytes by human T-cell leukemia virus type I X-region genes transduced by a Herpesvirus saimiri vector. Proc Natl Acad Sci U S A **86:**3351-5.

12. **Grossman, W. J., J. T. Kimata, F. H. Wong, M. Zutter, T. J. Ley, and L. Ratner.** 1995. Development of leukemia in mice transgenic for the tax gene of human T-cell leukemia virus type I. Proc Natl Acad Sci U S A **92:**1057-61.

13. **Harris, R. S., and M. T. Liddament.** 2004. Retroviral restriction by APOBEC proteins. Nat Rev Immunol **4:**868-77.

14. **Hasegawa, H., H. Sawa, M. J. Lewis, Y. Orba, N. Sheehy, Y. Yamamoto, T. Ichinohe, Y. Tsunetsugu-Yokota, H. Katano, H. Takahashi, J. Matsuda, T. Sata, T. Kurata, K. Nagashima, and W. W. Hall.** 2006. Thymus-derived leukemia-lymphoma in mice transgenic for the Tax gene of human T-lymphotropic virus type I. Nat Med **12:**466-72.

15. **Honjo, T., H. Nagaoka, R. Shinkura, and M. Muramatsu.** 2005. AID to overcome the limitations of genomic information. Nat Immunol **6:**655-61.

16. **Jin, D. Y., F. Spencer, and K. T. Jeang.** 1998. Human T cell leukemia virus type 1 oncoprotein Tax targets the human mitotic checkpoint protein MAD1. Cell **93:**81-91.

17. **Kannagi, M., S. Harada, I. Maruyama, H. Inoko, H. Igarashi, G. Kuwashima, S. Sato, M. Morita, M. Kidokoro, M. Sugimoto, and et al.** 1991. Predominant recognition of human T cell leukemia virus type I (HTLV-I) pX gene products by human CD8+ cytotoxic T cells directed against HTLV-I-infected cells. Int Immunol **3:**761-7.

18. **Koiwa, T., A. Hamano-Usami, T. Ishida, A. Okayama, K. Yamaguchi, S. Kamihira, and T. Watanabe.** 2002. 5'-long terminal repeat-selective CpG methylation of latent human T-cell leukemia virus type 1 provirus in vitro and in vivo. J Virol **76:**9389-97.

19. **Kuo, Y. L., and C. Z. Giam.** 2006. Activation of the anaphase promoting complex by HTLV-1 tax leads to senescence. EMBO J **25:**1741-52.

20. **Lairmore, M. D., L. Silverman, and L. Ratner.** 2005. Animal models for human T-lymphotropic virus type 1 (HTLV-1) infection and transformation. Oncogene **24:**6005-15.

21. **Liddament, M. T., W. L. Brown, A. J. Schumacher, and R. S. Harris.** 2004. APOBEC3F properties and hypermutation preferences indicate activity against HIV-1 in vivo. Curr Biol **14:**1385-91.

22. **Liu, B., S. Hong, Z. Tang, H. Yu, and C. Z. Giam.** 2005. HTLV-I Tax directly binds the Cdc20-associated anaphase-promoting complex and activates it ahead of schedule. Proc Natl Acad Sci U S A **102:**63-8.

23. **Liu, M., L. Yang, L. Zhang, B. Liu, R. Merling, Z. Xia, and C. Z. Giam.** 2008. Human T-cell leukemia virus type 1 infection leads to arrest in the G1 phase of the cell cycle. J Virol **82:**8442-55.

24. **Luo, K., B. Liu, Z. Xiao, Y. Yu, X. Yu, R. Gorelick, and X. F. Yu.** 2004. Amino-terminal region of the human immunodeficiency virus type 1 nucleocapsid is required for human APOBEC3G packaging. J Virol **78:**11841-52.

25. **Mahieux, R., R. Suspene, F. Delebecque, M. Henry, O. Schwartz, S. Wain-Hobson, and J. P. Vartanian.** 2005. Extensive editing of a small fraction of human T-cell leukemia virus type 1 genomes by four APOBEC3 cytidine deaminases. J Gen Virol **86:**2489-94.

26. **Malim, M. H., and M. Emerman.** 2008. HIV-1 accessory proteins--ensuring viral survival in a hostile environment. Cell Host Microbe **3:**388-98.

27. **Mansky, L. M.** 2000. In vivo analysis of human T-cell leukemia virus type 1 reverse transcription accuracy. J Virol **74:**9525-31.

28. **Matsumoto, Y., H. Marusawa, K. Kinoshita, Y. Endo, T. Kou, T. Morisawa, T. Azuma, I. M. Okazaki, T. Honjo, and T. Chiba.** 2007. Helicobacter pylori infection triggers aberrant expression of activation-induced cytidine deaminase in gastric epithelium. Nat Med **13:**470-6.

29. **Matsuoka, M., and K. T. Jeang.** 2007. Human T-cell leukaemia virus type 1 (HTLV-1) infectivity and cellular transformation. Nat Rev Cancer **7:**270-80.

30. **Mitchell, M. S., E. T. Bodine, S. Hill, G. Princler, P. Lloyd, H. Mitsuya, M. Matsuoka, and D. Derse.** 2007. Phenotypic and genotypic comparisons of human T-cell leukemia virus type 1 reverse transcriptases from infected T-cell lines and patient samples. J Virol **81:**4422-8.

31. **Miyazaki, M., J. Yasunaga, Y. Taniguchi, S. Tamiya, T. Nakahata, and M. Matsuoka.** 2007. Preferential Selection of Human T-Cell Leukemia Virus Type 1 Provirus Lacking the 5' Long Terminal Repeat during Oncogenesis. J Virol **81:**5714-23.

32. **Miyazato, P., J. Yasunaga, Y. Taniguchi, Y. Koyanagi, H. Mitsuya, and M. Matsuoka.** 2006. De novo human T-cell leukemia virus type 1 infection of human lymphocytes in NOD-SCID, common gamma-chain knockout mice. J Virol **80:**10683-91.

33. **Navarro, F., B. Bollman, H. Chen, R. Konig, Q. Yu, K. Chiles, and N. R. Landau.** 2005. Complementary function of the two catalytic domains of APOBEC3G. Virology **333:**374-86.

34. **Okazaki, S., R. Moriuchi, N. Yosizuka, K. Sugahara, T. Maeda, I. Jinnai, M. Tomonaga, S. Kamihira, and S. Katamine.** 2001. HTLV-1 proviruses encoding non-functional TAX in adult T-cell leukemia. Virus Genes **23:**123-35.

35. **Renjifo, B., I. Borrero, and M. Essex.** 1995. Tax mutation associated with tropical spastic paraparesis/human T-cell leukemia virus type I-associated myelopathy. J Virol **69:**2611-6.

36. **Saito, M., Y. Furukawa, R. Kubota, K. Usuku, S. Sonoda, S. Izumo, M. Osame, and M. Yoshida.** 1995. Frequent mutation in pX region of HTLV-1 is observed in HAM/TSP patients, but is not specifically associated with the central nervous system lesions. J Neurovirol **1:**286-94.

37. **Sasada, A., A. Takaori-Kondo, K. Shirakawa, M. Kobayashi, A. Abudu, M. Hishizawa, K. Imada, Y. Tanaka, and T. Uchiyama.** 2005. APOBEC3G targets human T-cell leukemia virus type 1. Retrovirology **2:**32.

38. **Satou, Y., J. Yasunaga, M. Yoshida, and M. Matsuoka.** 2006. HTLV-I basic leucine zipper factor gene mRNA supports proliferation of adult T cell leukemia cells. Proc Natl Acad Sci U S A **103:**720-5.

39. **Takatsuki, K.** 2005. Discovery of adult T-cell leukemia. Retrovirology **2:**16.

40. **Takeda, S., M. Maeda, S. Morikawa, Y. Taniguchi, J. Yasunaga, K. Nosaka, Y. Tanaka, and M. Matsuoka.** 2004. Genetic and epigenetic inactivation of tax gene in adult T-cell leukemia cells. Int J Cancer **109:**559-67.

41. **Tamiya, S., M. Matsuoka, K. Etoh, T. Watanabe, S. Kamihira, K. Yamaguchi, and K. Takatsuki.** 1996. Two types of defective human T-lymphotropic virus type I provirus in adult T-cell leukemia. Blood **88:**3065-73.

42. **Taniguchi, Y., K. Nosaka, J. Yasunaga, M. Maeda, N. Mueller, A. Okayama, and M. Matsuoka.** 2005. Silencing of human T-cell leukemia virus type I gene transcription by epigenetic mechanisms. Retrovirology **2:**64.

43. **Trono, D.** 2004. Retroviruses under editing crossfire: a second member of the human APOBEC3 family is a Vif-blockable innate antiretroviral factor. EMBO Rep **5:**679-80.

44. **Turelli, P., and D. Trono.** 2005. Editing at the crossroad of innate and adaptive immunity. Science **307:**1061-5.

45. **Wiegand, H. L., B. P. Doehle, H. P. Bogerd, and B. R. Cullen.** 2004. A second human antiretroviral factor, APOBEC3F, is suppressed by the HIV-1 and HIV-2 Vif proteins. EMBO J **23:**2451-8.

46. **Yoshida, M., Y. Satou, J. Yasunaga, J. Fujisawa, and M. Matsuoka.** 2008. Transcriptional control of spliced and unspliced human T-cell leukemia virus type 1 bZIP factor (HBZ) gene. J Virol **82:**9359-68.

47. **Yu, Q., R. Konig, S. Pillai, K. Chiles, M. Kearney, S. Palmer, D. Richman, J. M. Coffin, and N. R. Landau.** 2004. Single-strand specificity of APOBEC3G accounts for minus-strand deamination of the HIV genome. Nat Struct Mol Biol **11:**435-42.

FIGURE LEGENDS

Fig. 1. **Schema of the distribution of nonsense mutations in HTLV-1 provirus.** The *p13* gene has the same reading frame as the *p30* gene. A star indicates a nonsense mutation observed in an ATL case.

Fig. 2. Mutations in non-tumor cell-derived proviral sequence in ATL patients. HTLV-1 genomes from two ATL samples with internal deleted regions were schematically demonstrated. As indicated by horizontal arrows, one of the PCR primers was set in the deleted region in order to specifically amplify a proviral region originated from non-tumor cells (See Materials & Methods). Nested PCR products, represented by short lines, were subject to direct sequencing. Nucleotide bases mutated in tumor cells according to the criteria mentioned in Materials & Methods were displayed. A base substitution is considered as a polymorphism if the same base was found also in the non-tumor cells derived from the same patient sample. Positions underlined are those at which a G-to-A mutation will result in a nonsense mutation.

Fig. 3. Target sequence preference of G-to-A mutations. (A) Sequence context of G-to-A mutations. In all G-to-A mutations, the percentage of ones at the indicated dinucleotide sequence context is shown (left panel). In GG context, the influence of 5' (middle panel) and 3' (right panel) nucleotides neighboring the GG dinucleotide on G-to-A mutations is displayed. **(B)** Frequency of G-to-A mutations on sites of tetranucleotides containing a central GG. HTLV-1 genome sequence from ATL-25 was

used as a standard sequence with mutated bases corrected to calculate the total number of GG sites that exist in the provirus genome (See Materials & Methods). The percentage of GG sites that underwent G-to-A mutation for at least once in 60 ATL cases was plotted against each sequence context. Mutated bases are underlined. CGGG and TGGG were the preferred targets for G-to-A mutation in the HTLV-1 provirus. A C at the +2 position (NGGC) was disfavored.

Fig. 4. Expression of the *hA3G* and *AID* gene in ATL cases and in HTLV-1 associated cell lines. Expression of the *hA3G* (A) and *AID* gene (B) was investigated by real-time PCR. Normal resting T cells isolated from 3 healthy blood donors were used as a control. *ACTN* was used as an internal control. Relative quantification was performed using a comparative Ct method (Applied Biosystems). Expression of *hA3G* and *AID* in resting T cells was artificially set as 1. Number of nucleotide mutations in the 7 ATL cases was shown below. Among mutated sites, number of AID preferred sites was demonstrated in parenthesis. (C) Number of nucleotide mutations in the 7 ATL cases. Among mutated sites, the number of AID preferred sites was demonstrated in parenthesis.

Fig. 5. In vitro editing of HTLV-1 genome by hA3G. (A) Mutation matrix for hyperedited *tax* gene sequences derived from cloned 86˚C 3DPCR products. n indicates the number of bases sequenced. (B) A selection of hypermutated HTLV-1 *tax* sequences

from tissue culture. The sequence is given with respect to the viral plus strand. Only differences are shown. # and % denotes respectively the number of mutations and the percentage of G-to-A transitions per sequence. The base composition of the sequence is 75 T (24 %), 113 C (36.2 %), 60 G (19.2 %) and 64 A (20.5 %). Position number of the first base is indicated above the base. Arrows indicate positions where G-to-A-mediated nonsense mutations were detected in ATL cases. (C) Bulk dinucleotide context of HTLV tax region DNA strand by hA3G. The ordinate represents the substitution frequency as a function of the 3' nucleotide. Mutated G residues are underlined. A χ^2 analysis showed that the observed frequencies by hA3G deviated significantly from the expected values ($p < 0.001$).

1

Table 1. Abortive genetic changes of HTLV-1 viral genes in 60 ATL cases[a]

	gag	pol	env	p12	p30	p13	p27	tax	HBZ
ATL-2		DEL							
ATL-3								W248*	
ATL-7			DEL						
ATL-8			DEL				W139*		
ATL-10	DEL	DEL	DEL						
ATL-11		W680*	W427*						
ATL-12	DEL		W387*		W165*	W11*	W139*	W56*	
ATL-13	W59*	W680*	DEL		W165*	W11*		W56*	
					W190*	W36*			
ATL-14							DEL	DEL	
ATL-15		W680*							
ATL-27		DEL							
ATL-29			W431*		W165*	W11*		W56*	
ATL-32		W271*							
ATL-36								DEL	
ATL-37		W485*	W50*					W56*	
		W680*							
ATL-39				W87*					
ATL-40					R184*	R30*			
ATL-41		W485*						W56*	
ATL-42			DEL						
ATL-43		DEL							
ATL-44	DEL	DEL		W87*					
ATL-45	DEL							Y286*	
ATL-46	DEL	DEL							
ATL-47	DEL	DEL							
ATL-48	DEL							IN	
ATL-49	DEL	DEL							
ATL-51					W178*	W24*			
ATL-53	DEL								

[a]Only cases with deletion, insertion or non-sense mutations are shown. DEL, deletion. IN, insertion. W, Tryptophan. *, stop codon. Number indicates position of non-sense mutation.

Table 2. Concordance rate of mutations between leukemic cells and non-leukemic cells in five ATL cases.

	Total[a]	G-A	C-T	A-G	T-C	A-C	A-T	C-G	C-A	T-G	G-C	G-T	T-A
ATL-8	10 (4)	2 (1)	3 (1)	2 (0)	1 (0)	-	-	1 (1)	-	-	1 (1)	-	-
ATL-12	25 (8)	8 (6)	7 (2)	5 (0)	3 (0)	-	-	2 (0)	-	-	-	-	-
ATL-13	36 (20)	10 (8)	8 (3)	7 (2)	2 (1)	2 (0)	-	-	1 (0)	1 (0)	3 (3)	1 (0)	1 (0)
ATL-14	12 (2)	1 (1)	7 (0)	1 (0)	2 (0)	-	-	-	1 (1)	-	-	-	-
ATL-45	16 (3)	3 (1)	3 (0)	4 (0)	2 (1)	-	3 (0)	-	1 (1)	-	-	-	-
Total	99 (37)	24 (17)	28 (6)	19 (2)	10 (2)	2 (0)	3 (0)	3 (1)	3 (2)	1 (0)	4 (4)	1 (0)	1 (0)
Concordance rate (%)	37%	71%	21%	11%	20%	0%	0%	33%	67%	0%	100%	0%	0%

Each entry *A(B)* shows the number *A* of each type of mutation in ATL judged by our criteria and the number *B* of confirmed mutations. Concordance rate of mutations was calculated by dividing *B* by *A*.

[a]Number of mutations of all types in each ATL case.

Table 3. Distribution of hA3G target sequence in the plus strand of HTLV-1 genome.

Genome Regions	5'LTR	*gag*	*pol*	*env*	*p27*	*p30*	*tax*	*HBZ*	3'LTR
Size (bp)	755	1290	2691	1467	570	726	1062	621	755
CGGX[a]	8	12	16	8	3	2	5	1	7
CGGX/1 kb	10.60	9.3	5.95	5.45	5.26	2.75	4.71	1.61	9.27
TGGX	3	10	25	17	10	7	11	1	3
TGGX/1 kb	3.97	7.75	9.29	11.59	17.54	9.64	10.36	1.61	3.97

[a]X indicates A, T or G.

Table 4. Mutations in the *pol* region of 10 HTLV-1 carriers[a]

	Number and type of single nucleotide substitution									Sequence context of G-to-A mutations			
	G-A	C-T	A-G	T-C	A-C	C-G	G-C	G-T	T-A	GA	GC	GT	GG
C72 (25)	5	1					1						5
C79 (20)	8		1		1			1		1	1		6
C63 (28)	1							1				1	
C40 (26)	13		1										13
C32 (22)		2											
C62 (22)	1	2											1
C87 (22)	14									1	1		13
C5 (23)	19	1						1	2				19
C29 (25)	9					1		1			1		8
C82 (27)	7			1							1		6
Total	77	6	2	1	1	1	1	4	2	2	4	1	71

[a]Number of clones studied for each samples was indicated in the brackets. Mutated bases are underlined.

Table 5. Mutations in the *tax* region of 10 HTLV-1 carriers[a]

	Number and type of single nucleotide substitution								Sequence context of G-to-A mutations			
	G-A	C-T	A-G	A-C	C-A	G-C	G-T	T-A	GA	GC	GT	GG
C72 (24)	3								1			2
C79 (24)	2	4		2	1	1					2	
C63 (21)	19											19
C40 (21)	2											2
C32 (25)	4	1										4
C62 (27)	5	1								1		4
C87 (24)	1	1							1			
C5 (26)	2											2
C29 (25)	3		3					1				3
C82 (22)	5						1					5
Total	46	7	3	2	1	1	1	1	2	1	2	41

[a]Number of clones studied for each samples was indicated in the brackets. Mutated bases are underlined.

Table 6. Nonsense mutations in ten HTLV-1 carriers.

Genes	**pol**		**env**						**p12**	**p30/p13**	**p27**[a]		**tax**		**HBZ**
Position	**3955**	**4540**	**5465**	**6158**	**6342**	**6362**	**6482**	**6494**	**-**	**7361**	**7652**	**7660**	**7469**	**8045**	**-**
C72	**1/25**	-/25	-/23	-/23	-/23	-/23	-/23	-/23	-/31	-/31	-/31	-/31	**1/24**	-/24	-/31
C79	**2/20**	**2/20**	-/20	-/20	-/20	**1/20**	-/20	-/20	-/24	-/24	-/24	-/24	**1/24**	-/24	-/24
C63	-/28	-/28	-/20	-/20	-/20	-/20	-/20	-/20	-/24	**2/24**	-/24	-/24	**9/21**	-/21	-/24
C40	-/26	**13/26**	-/22	-/22	-/22	-/22	-/22	-/22	-/30	-/30	-/30	-/30	**2/21**	-/21	-/30
C32	-/22	-/22	-/22	-/22	-/22	-/22	-/22	-/22	-/32	-/32	-/32	**1/32**	-/25	-/25	-/32
C62	**1/22**	-/22	-/25	-/25	**2/25**	-/25	-/25	-/25	-/28	**1/28**	**1/28**		**1/27**	-/27	-/28
C87	**1/22**	**3/22**	-/20	-/20	-/20	-/20	-/20	-/20	-/21	-/21	**1/21**	-/21	-/24	-/24	-/21
C5	-/22	**19/22**	-/20	-/20	-/20	**1/20**	-/20	-/20	-/24	-/24	-/24	**1/24**	**1/26**	-/26	-/24
C29	-/25	**3/25**	-/20	**1/20**	-/20	-/20	-/20	**1/20**	-/24	**2/24**	**5/24**	-/24	**1/25**	-/25	-/24
C82	**3/27**	**2/27**	**1/20**	-/20	-/20	-/20	**1/20**	-/20	-/24	1/24	-/24	-/24	**1/22**	**1/22**	-/24

Each entry *A/B* in the table shows the number *A* of clones harboring nonsense mutations at a given nucleotide position, from a given carrier, along with the number of clones *B* analyzed in that region from that carrier.

[a]Sequences of exon 2 were determined.

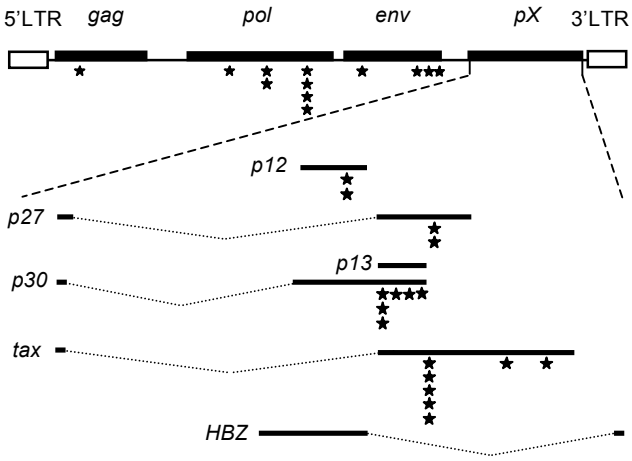


Figure 1

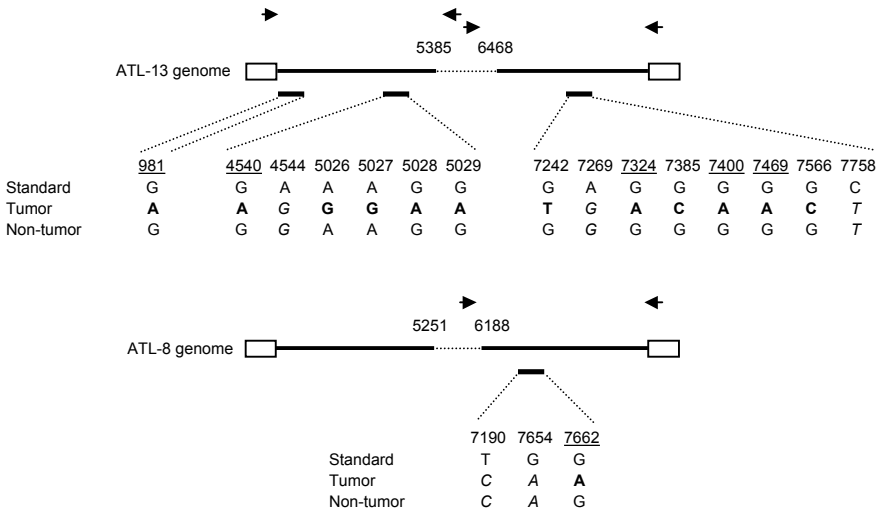
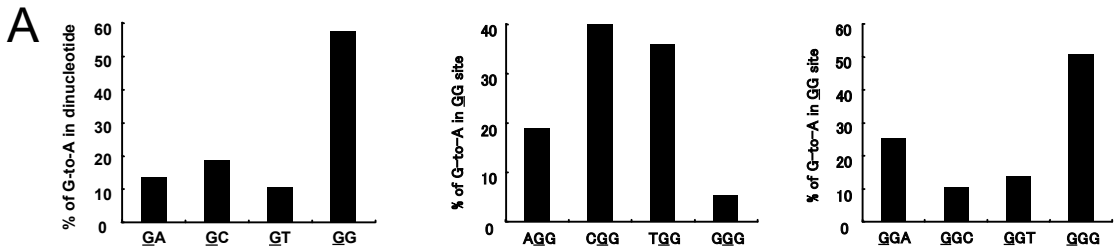
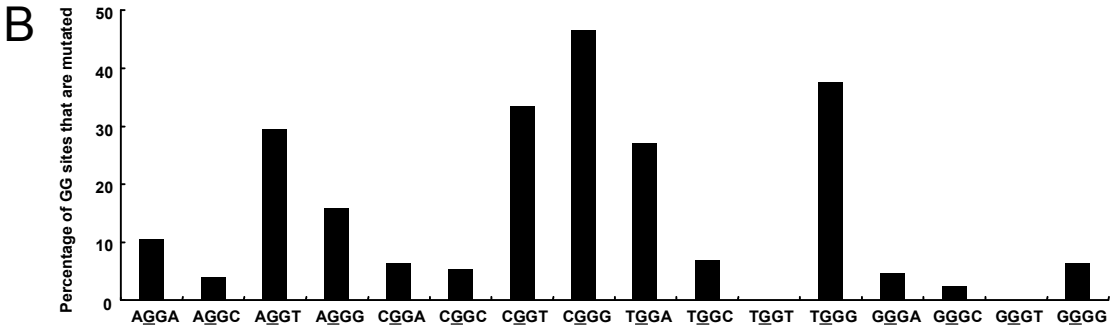


Figure 2

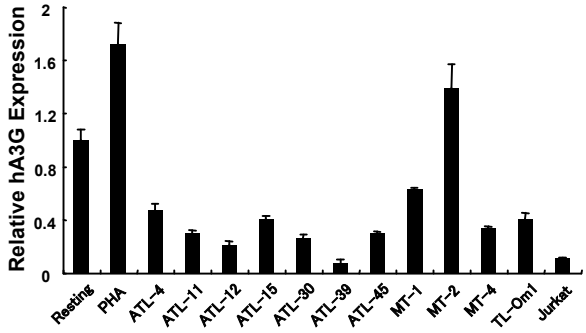




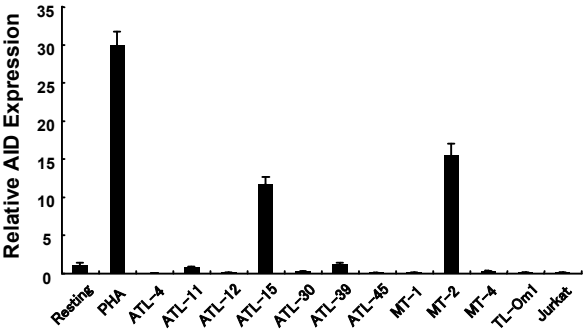
Site	AGGA	AGGC	AGGT	AGGG	CGGA	CGGC	CGGT	CGGG	TGGA	TGGC	TGGT	TGGG	GGGA	GGGC	GGGT	GGGG
Mutated	3	2	5	3	1	2	5	13	10	3	0	9	1	1	0	2
All	29	51	17	19	16	38	15	28	37	44	10	24	22	42	8	32
%	10.3	3.9	29.4	15.8	6.3	5.3	33.3	46.4	27.0	6.8	0.0	37.5	4.5	2.4	0.0	6.3

Figure 3

A



B



C

	G-A	C-T	A-G	T-C	A-C	A-T	C-A	C-G	T-G	Total
ATL-4	3(0)	4(1)	1			2				10
ATL-11	6(1)	9(1)	3	8		1				27
ATL-12	8(2)	8(1)	4	3				2		25
ATL-15	10(3)	6(2)	2	2	1	1			1	23
ATL-30	2(0)	1(0)				1				4
ATL-39	2(0)	8(0)	2	5	3	1				21
ATL-45	3(0)	3(0)	4	2		3	1			16

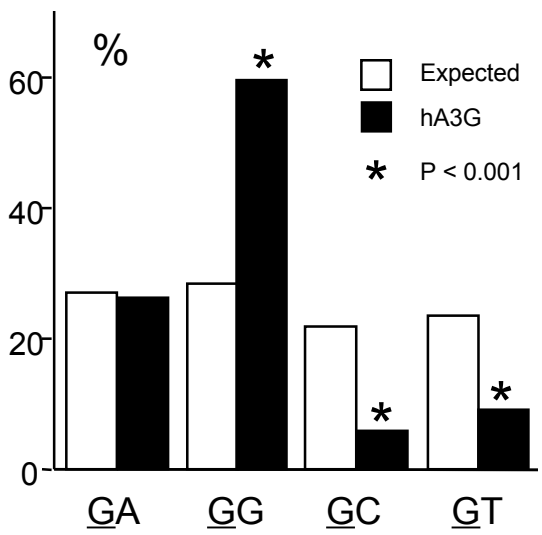
Figure 4

A

	T	C	G	A
T		2	0	0
C	0		0	1
G	0	0		231
A	0	0	1	

HTLV+A3G
n=3432 bp

C



B

```
7291
CCATGCTTATTATCAGCCCACTTCCCAGGGTTTGGACAGAGTCTTCTTTTCGGATACCCAGTCTACGTGTTTGGAGACTGTGTACAAGGCGATTGGTGCCCCAT   #   %
.............................A.....A....A.....................................................................   8  13
.................................A................A.................A.............A.............A.............  12  20
.................................A................A.....................................A......A.........  19  32
............................AA...A....A...........A.............A.....................A......AA........  26  43
............................AAA...A....A...........AA............A.....AA.A...A.A.....A..A...A.........  29  48

CTCTGGGGGACTATGTTCGGCCCGCCTACATCGTCACGCCCTACTGGCCACCTGTCCAGAGCATCAGATCACCTGGGACCCCATCGATGGACGCGTTATCAGCT
....A.....................................A............A.............................A.................
....AAAA..................................A............A.............A............A.................
....AAAA.........A........................A.......A....A.A.....A.......AA............A.................
....AAAAA.....A...A.....................A.....................................AA.........A..A...A.A.........
....AAAAA.........A.....................................................AA.........A..AA..A...............

CAGCTCTACAGTTCCTTATCCCTCGACTCCCCTCCTTCCCCACCCAGAGAACCTCTAAGACCCTCAAGGTCCTTACCCCGCCAATCACTCATACAACCCCCAAC
.........................A.....................................G.....................................................
.....................................................................................................................
.................................................A.................A.....................................
.........................A.....................A.................A.....................................
.................................................A.....................................................
```

Figure 5